Cytoscape Tutorial (in progress)

From STAMPS

Contents

Goals

1) To import and build a network from USC Microbial Observatory at the San Pedro Ocean Time Series (SPOT),

from Automated Ribosomal Intergenic Spacer Analysis (ARISA) data.

2) To create smaller networks from the full network by

a) manually selecting a few nodes,
b) filtering based on node features (ie, taxonomic assignment) and then creating a network.

3) To edit network vizualization by changing node colors/shapes/size and edge colors.
4) To import node attributes onto network.

The data we are working with today all come from samples taken from the San Pedro Ocean Time Series (SPOT) offshore of Los Angeles California in the Pacific Ocean. Samples were collected from the chlorophyl maximum layer (the depth at which chlorophyll is the most concentrated, usually around 15 to 100m below the surface. All samples were collected between 2000 and 2003 from the Chlorophyl Maximum Depth.

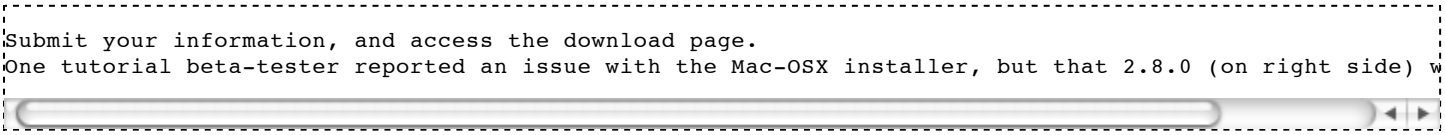
More background info on SPOT is available here: http://dornsife.usc.edu/labs/usc-microbial-observatory/

In this example, nodes are ARISA OTUs (bacteria), 18S TRFLP OTUs (eukaryotic protists) or environmental parameters. Edges are the connections between nodes, and are labelled with the local similarity analysis values (LSA, a time-shifted-based correlation coefficient). These edge values are generated by the LSA program and can be directional. It is important to note that any type of correlation matrix can be visualized in Cytoscape - co-expression data, Pearsons, Spearman's, etc.

Software and Test Data

To run Cytoscape

1. Download Cytoscape Here (http://www.cytoscape.org/download.html)



2. Run installer.
3. **Download** this Data for LSA (right-click, Save Link As)
*Be sure to download **both** data files*. The first file is the edge attributes file; the second file is the node attributes file.
Edge Data for LSA Media:EdgesSteeleSTAMPScc.txt
Node Data for LSA Media:NodesSteeleSTAMPScc.txt
4. Open Cytoscape

Data Input Files Description

EdgesSteeleSTAMPScc.txt lists the edge connections and attributes (positive or negative, values, etc.).
NodesSteeleSTAMPScc.txt lists the nodes and various attributes.

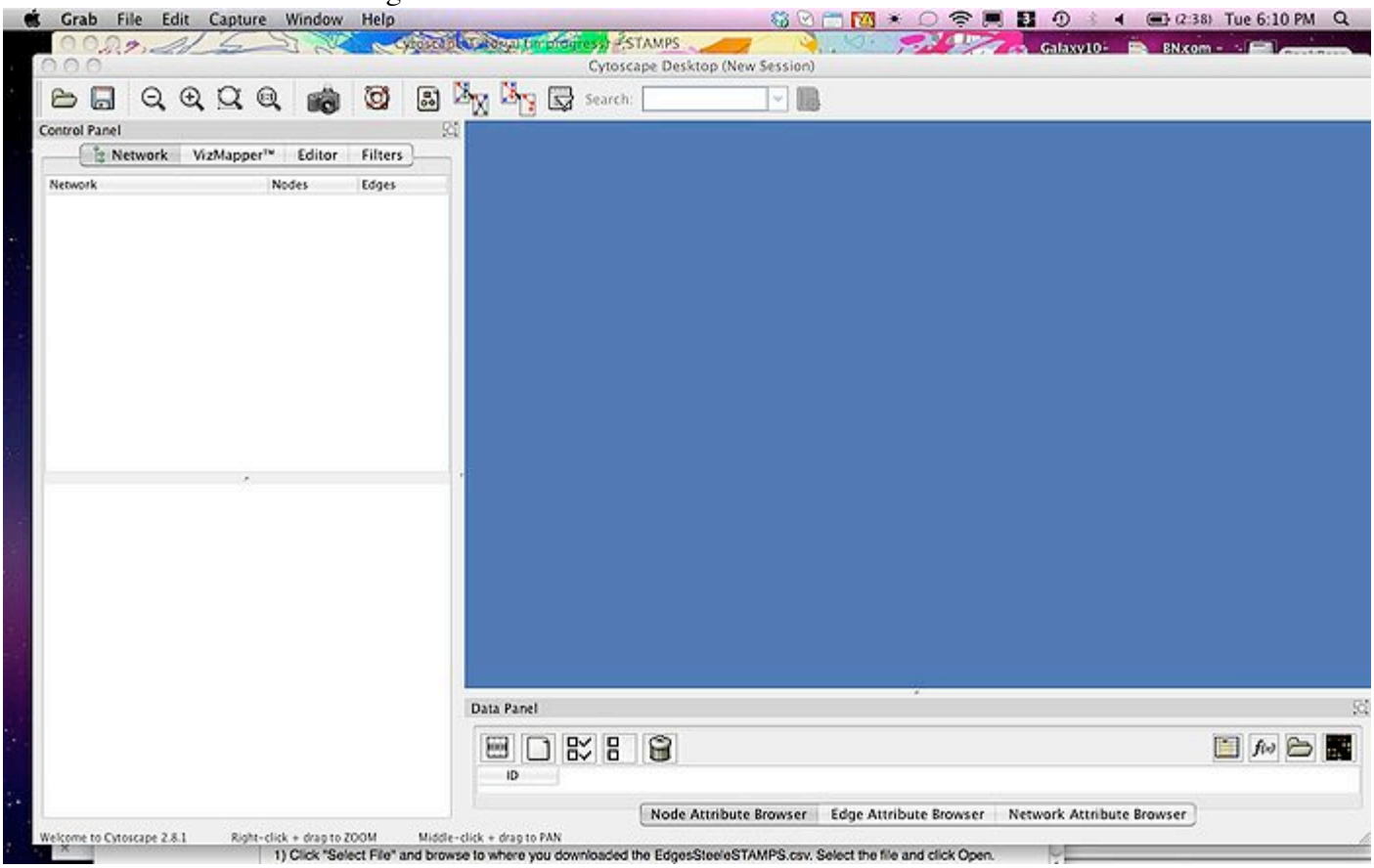
Orientation to Cytoscape

First, find your mouse/trackpad/multi-touch trackpad, etc. Remember right and left-click? :)

When you open the program you should see:
Control Panel (left panel)

Option: (click on triangles to see the other options)

Network: lists the networks as you import or create them
VizMapper: visual properties of the network (ie, the look) can be adjusted here
Editor: to add individual edges or nodes.
Filters: to select subsets of edges or nodes.



How to import a network from the LSA result files

From menu bar:

```
File > Import > Network from Table (Text/MS Excel).
Click "Select File" and browse to where you downloaded the EdgesSteeleSTAMPScc.txt
Double click on the file (or click > open).  Do not click Import yet!  Follow the instructions in the n
```

Note: This file includes connections with all p-values. It was parsed in Excel by a simple sort (and delete) of the LSA output.

1) Under Advanced box,

```
a. Select "Show text file import options". An additional set of options should appear.
b. Under Attribute Names: Select "Transfer first line as attribute names"
c. Leave rest of this panel as defaults. Still do not click import!  Follow instructions in the next bo
```

2) In the Interaction Definition (second box from top)

```
a. Source -> Column 1 (corresponds to X, ID of X Node)

b. Interaction Type -> Column 2 (corresponds to interaction type)

c. Target -> Column 3 (corresponds to Y, ID of Y Node) Do not click Import yet!  Follow the instructio
```

3) In Preview Window:

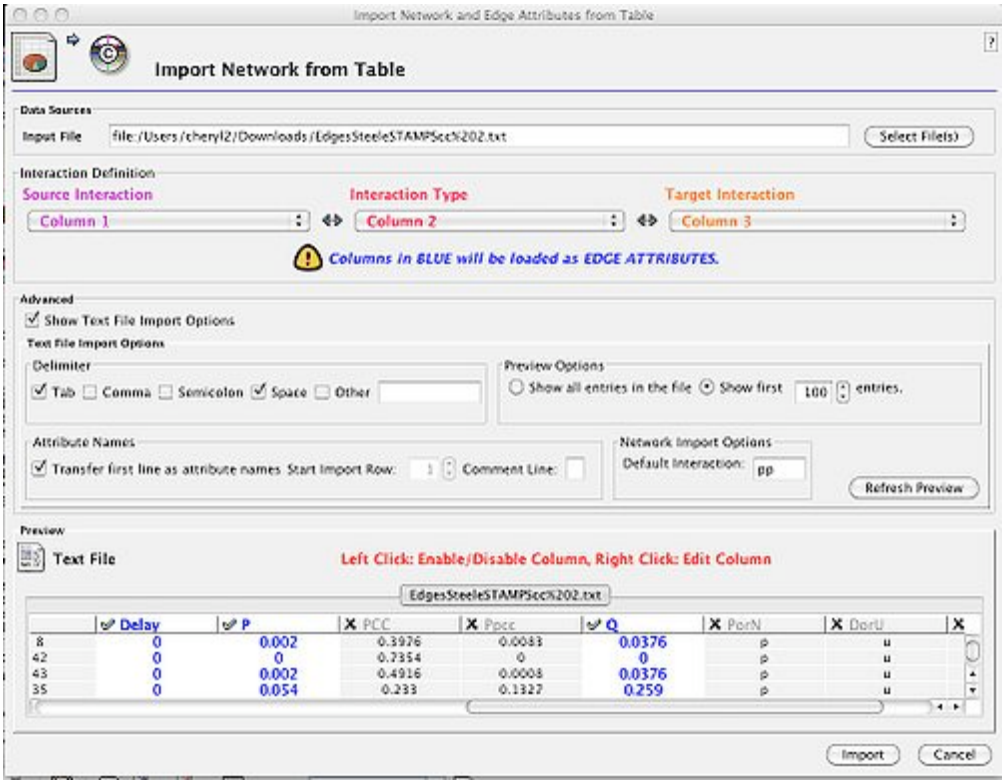
```
Click on column headings for (scroll to the right to find): LS, Delay, P, and Q.
```

Note: Verify that the columns are highlighted in blue. If blue and a checkmark, it will be imported. If grey and an "X", it will be ignored.

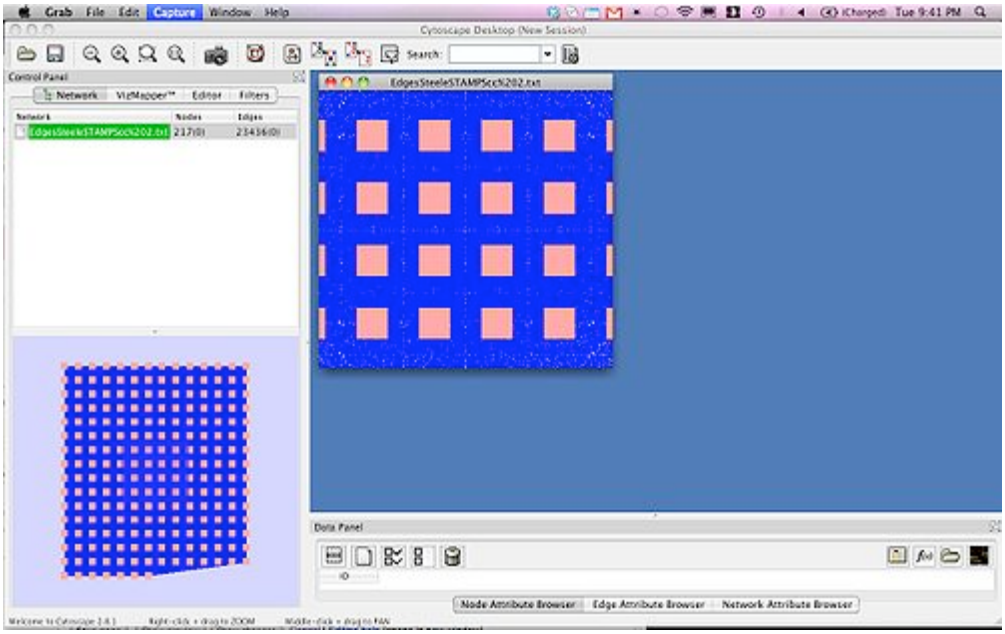
a. Delay (0,1,-1) indicates if the LSA correlation is based on any time shifts of 1 month before or after. Delay of 0 is no time correction.
b. LS value is the LSA result.
c. P indicates the p-value of the individual LSA value.

Nodes must not begin with a number or have spaces or other funny characters ($#!@...) although _ are ok. By importing additional node attributes later, you can re-label the nodes in the visualized network as you wish.

5) Click: Import (Finally). A new window confirming the import will show; click close.



Result: Network in square grid with all nodes and lots of lines connected. Number of nodes and edges imported in shown in the "Network panel".
Nodes are the pink circles, and LSA correlations (edges) are the blue lines. There are 23,000+ edges in the network -- many of these are not statistically significant. **Don't worry if it seems overwhelming! We will remove the non-significant LSA correlations in the upcoming filtering section.**



Manipulating Your Network

this section from the Cytoscape Open tutorial (http://opentutorials.cgl.ucsf.edu/index.php/Tutorial:Introduction_to_Cytoscape)

Now that you have a network loaded, you can interact with it in a number of ways:

- Start by clicking on the node at the upper left corner of the network. The node with turn

yellow. If you hold your mouse down over the node and drag it around the node will move on the screen.

- Now add another node to the selection by holding down the Shift key and clicking on a node. Note that both nodes are now selected (yellow). Again, move the nodes around. Node that both nodes will move.
- To select a group of nodes, hold the mouse down in the upper left-hand corner and drag your mouse over a region of the network. Again, a group of nodes will be selected and can be moved around on the screen.
- To zoom in on the selected nodes, click on the magnifying glass icon with a (+).
- To move the window around (change your viewing area), drag the small window outlined in blue around in the Network Overview Pane (lower left panel).
- Finally, zoom your network out by clicking on the magnifying glass icon with a (-). Also try the other two.

While useful, hand selecting nodes in dense networks can be error-prone and difficult. However, you can specifically search for a node by name or attribute:

- In the **Search:** box at the top of the screen, type in "f72". This will select that node and zoom the display to focus on it.

SAVE your network so far. As with anything, SAVE often.

```
File > Save > STAMPS_Cytoscape.cys
```

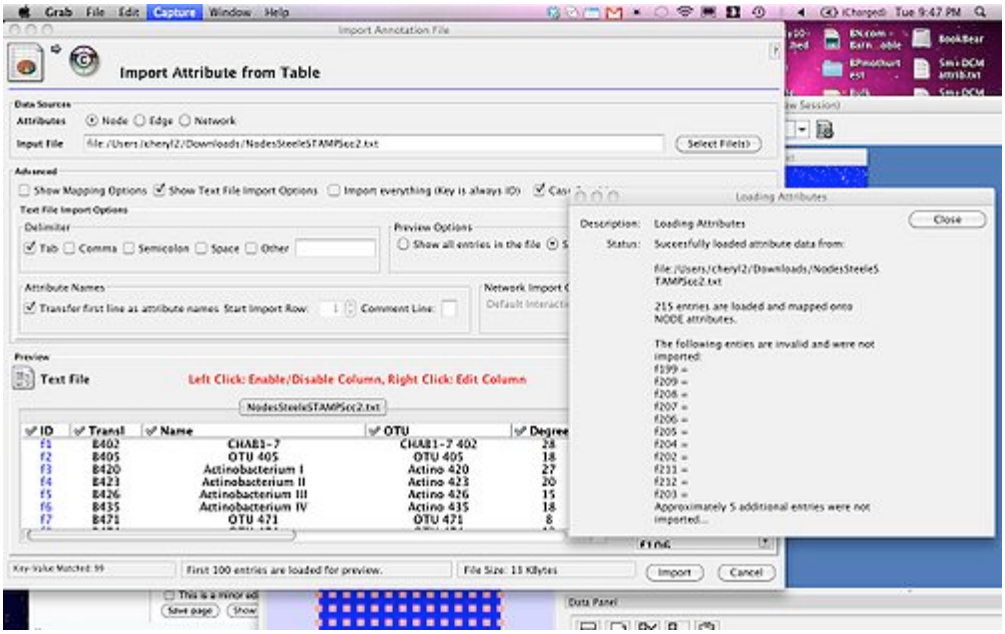
Importing Node Attributes

Note: These attributes can be any user-defined information, such as average abundance of an OTU or taxonomy.

1) From menubar, File > Import > Attribute from Table (Text/MS Excell).
2) Click "Node".
3) Click "Select File" and browse to where you downloaded the NodesSteeleSTAMPScc.txt. Select the file and click Open.
4) Under Advanced box,

```
Select "Show text file import options"
Under Attribute Names: Select "Transfer first line as attribute names".
```

5) Verify that all the input columns have a checkmark in the header and that the ID column is in blue text.
6) Click Import.
Note: An error message may appear about "null" values, but it's ok. We don't have values for all nodes for all attribute columns because they were not applicable.



To view the attribute info:

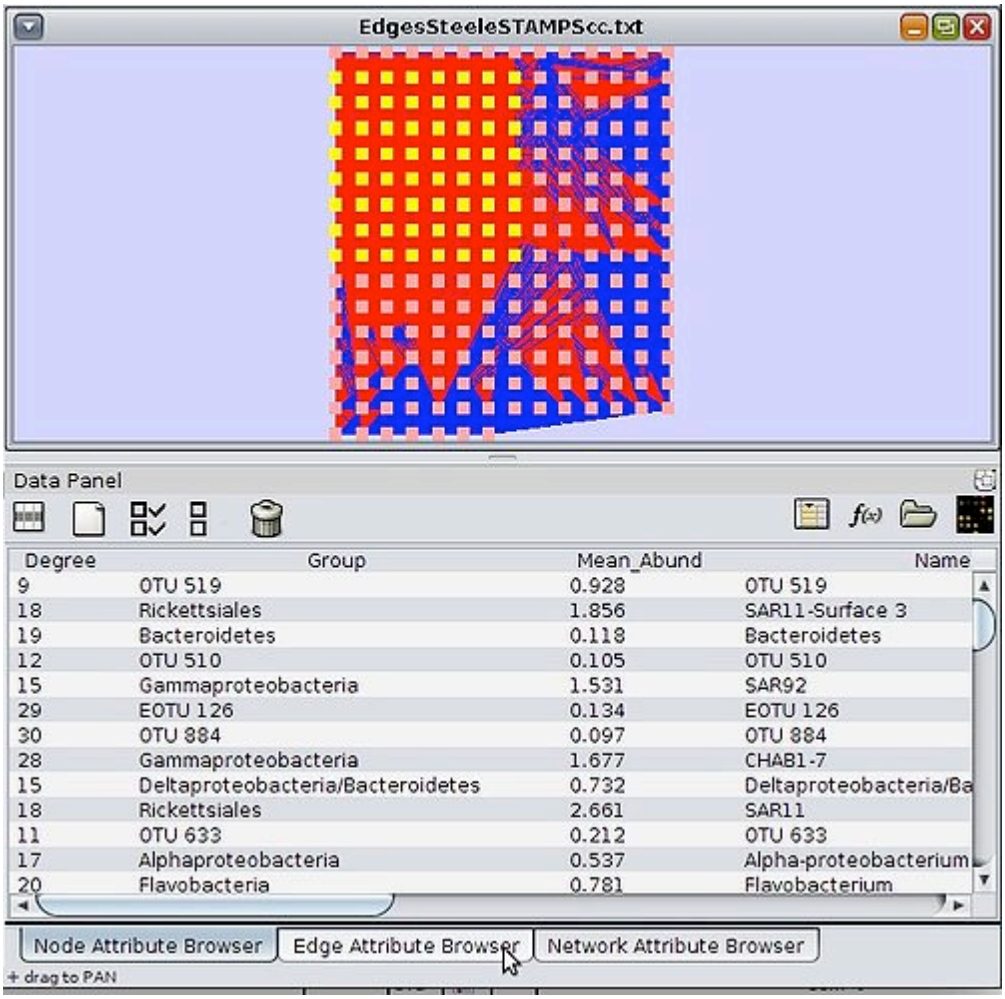
Click on third icon in the data panel (boxes with checks), to **Select all attributes**
It's located below network viewing area, near top of Data Panel (probably about halfway down in the middle of your screen).

Additional data columns should appear in the data panel window.

```
Select some nodes in the network window by shift-click.
```

Attribute data should appear in the data panel window. Only highlighted nodes (shown in yellow in the network viewer) will appear in the data panel. If you don't see anything, make sure you are in the "Node Attribute Browser". You can switch by clicking on the toolbar.

```
Look at edge attributes by clicking over to the "edge attribute browser."
Select some edges to see details (shift-click on a blue line).
```



Setting the Graphical Mapping Parameters: VizMapper

Mapping Methods

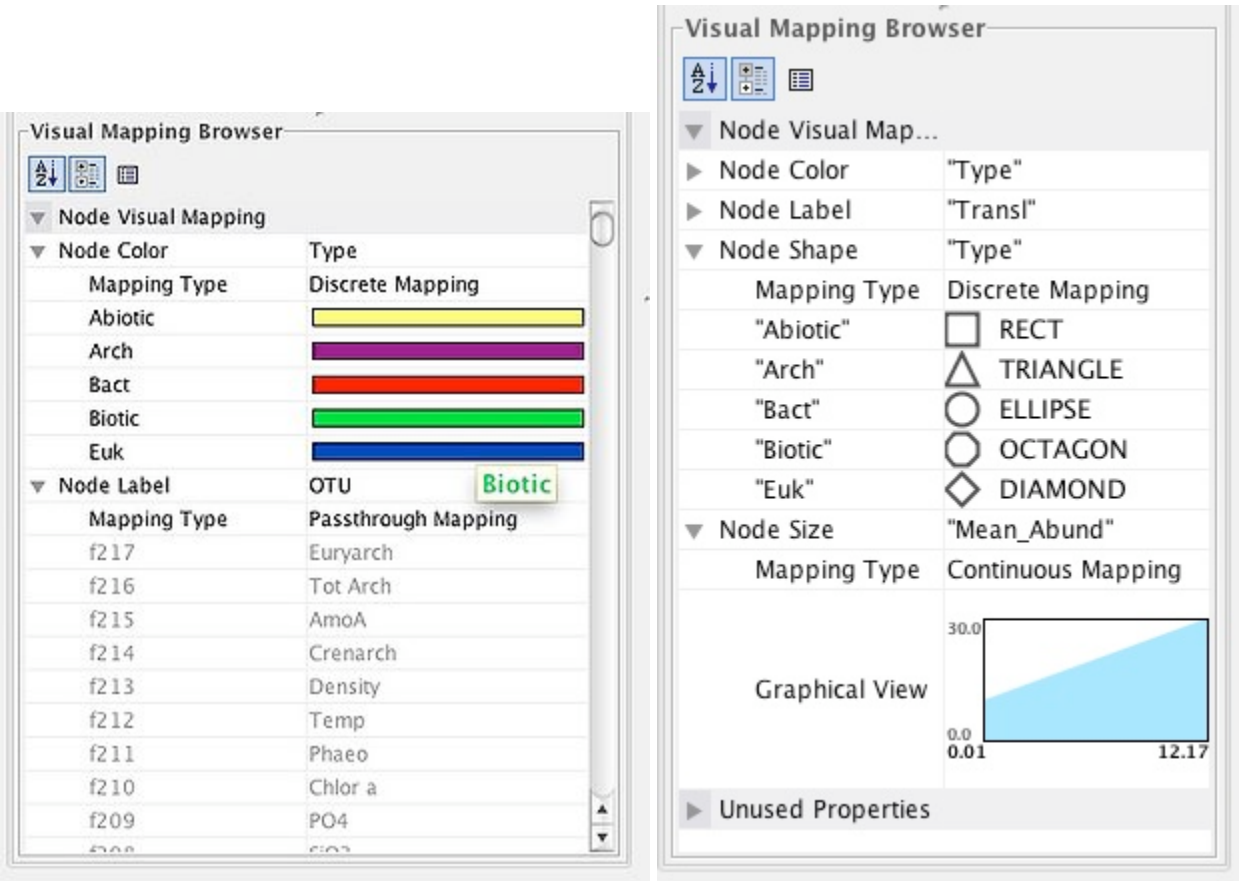
1) Discrete: Categorical but can be used with any input data type.
2) Pass-through: Attribute information is directly transferred to the node or edge.
3) Continuous: Numerical data is given a continuous range from the input values for size. Or for colors, a gradient will be created.

Nodes

```
On left-hand control panel, click over to VizMapper.
Under "Current Visual Style", click on tools symbol to the right of "Default" and "Create new style".
Enter a name.
```

For each of the following categories, double-click on the category name. Then select the desired mapping type and visual qualities.

```
Node Label: Set to "OTU". Mapping Type = Passthrough.

Node Color: Set to "Type". Mapping Type = Discrete. Pick a color for each of the categories listed by o
 the white space and then the three dots. A color palette window should open automatically.

Node Shape: Select "Type". Mapping Type = Discrete. Select shapes for each of the categories.

Node Size: Select "Mean_Abund". Mapping Type = Continuous. A window with a gradient should appear to ad
range of plotted sizes (or leave as is). Nodes should automatically change in size.
```

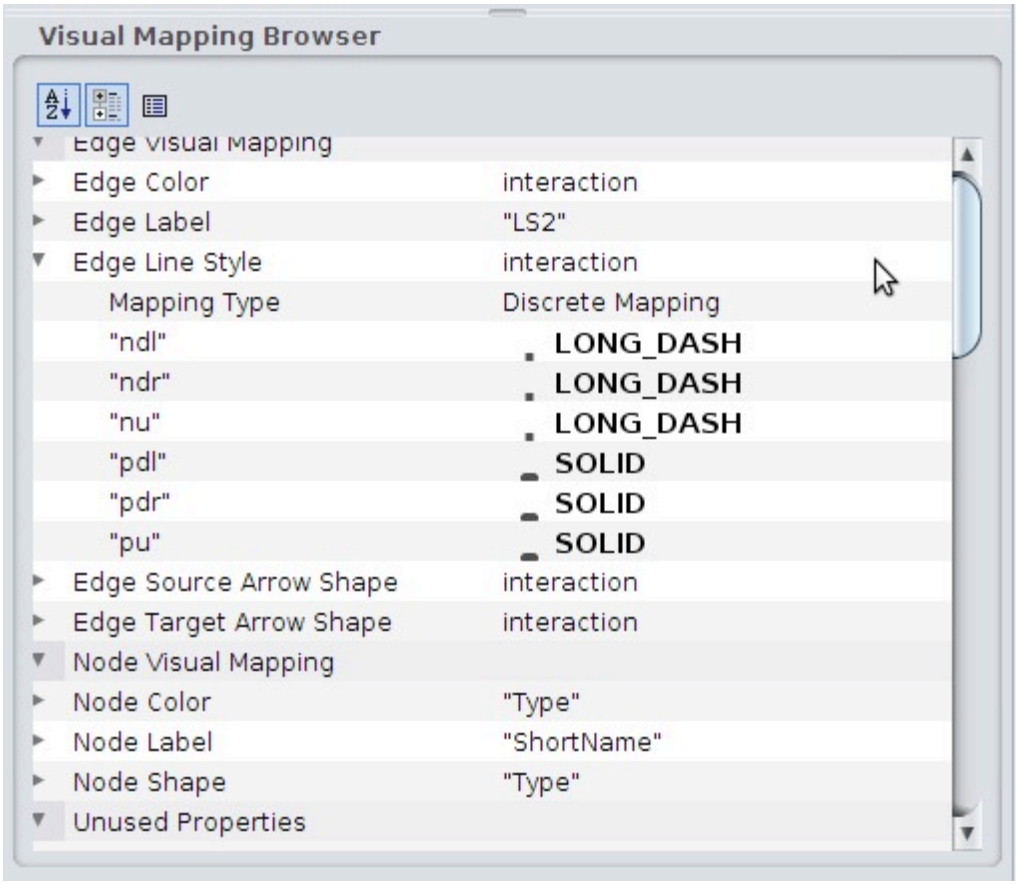


Edges

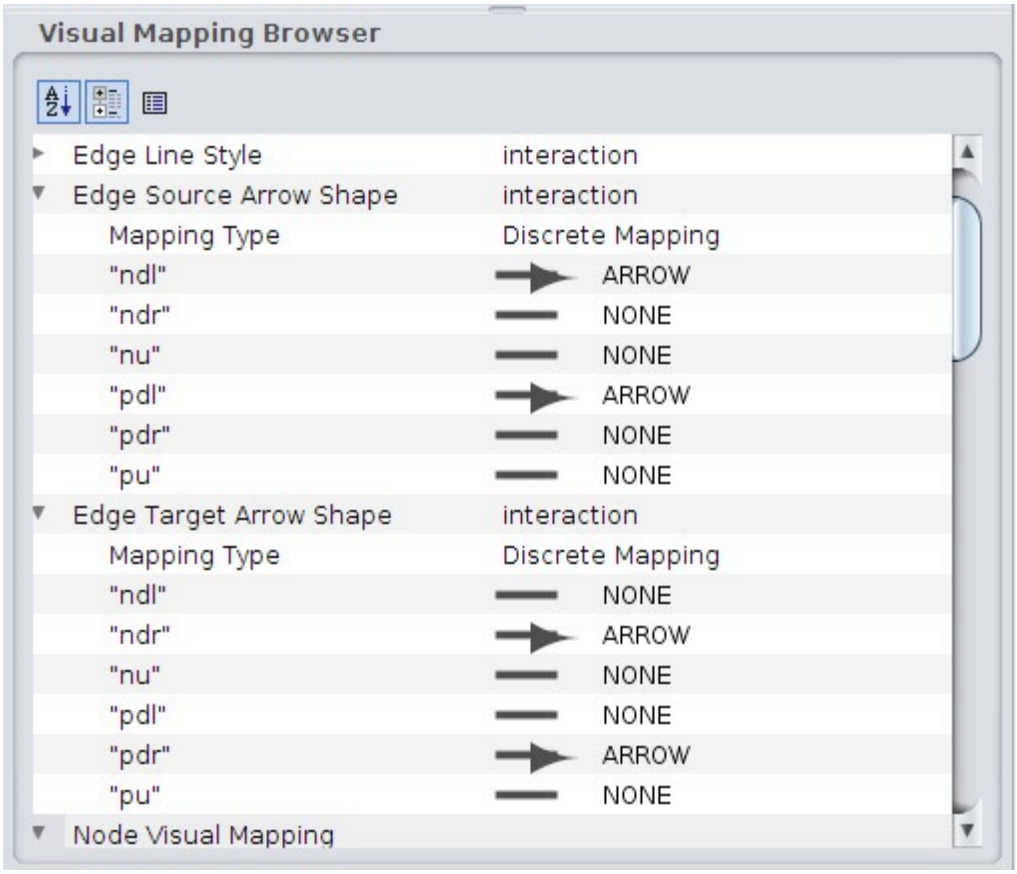
Edges are categorized into six types: X and Y nodes are determined from original node import table

ndl - negative correlation, directed left towards lagging month, so X follows Y
ndr - negative correlation, directed right towards lagging month, so Y follows X
nu - negative correlation, no time delay
pdl - positive correlation, directed left towards lagging month, so X follows Y
pdr - positive correlation, directed right towards lagging month, so Y follows X
pu - positive correlation, no time delay

```
Edge Label: LS, Passthrough

Edge Line Style: Interaction, Discrete
 Select solid lines for positive LSA values, and dashed for negative LSA values.
```




```
Edge Source and Target Arrow Shape: as shown below
*If you accidentally click the wrong Edge Source..., you can right-click on the "Edge Source ..."
label and select "Delete Mapping". This will return the category to your Unused Properties list.
```



Import & Export VizMap Properties and Legend

Create a .props file that you can use for future networks.

```
File>Export>VizMap Property File. Enter name, save to your directory.
```

Import your saved VizMapper settings (for future reference, when you make a new network).

```
File>Import> VizMap Property File.

Navigate to your VizMapper in the pull-down menu in VizMapper panel, under "Current Visual Style".
```

Export Legend

```
1) Click on tools (options) symbol under (or beside) "Current Visual Style".

2) Select "Create legend from current Visual Style."

3) Export (scroll down to bottom) and Save as graphics file (gif).
```

Result

Note: File is one image (vertical) and can be used a legend graphic for posters or other documents.

Filtering: Creating networks from networks

First Round Filtering for Only Statistically Significant Interactions

So now we have nodes (environmental parameters) connected by many interactions (LSA correlations). Many of these correlations are not statistically significant. Lets begin by making a new network that only has statistically significant correlations.
As a bit of background a Q value is a measure of significance that takes into account the fact that we are running *many* comparisons. Here, we will filter so that we only keep edges that represent correlations with a Q value better (lower) than 0.05 (5%). This would mean, essentially, that for this network, five percent of the putative correlations in this network are due to random chance.

To filter by Q value

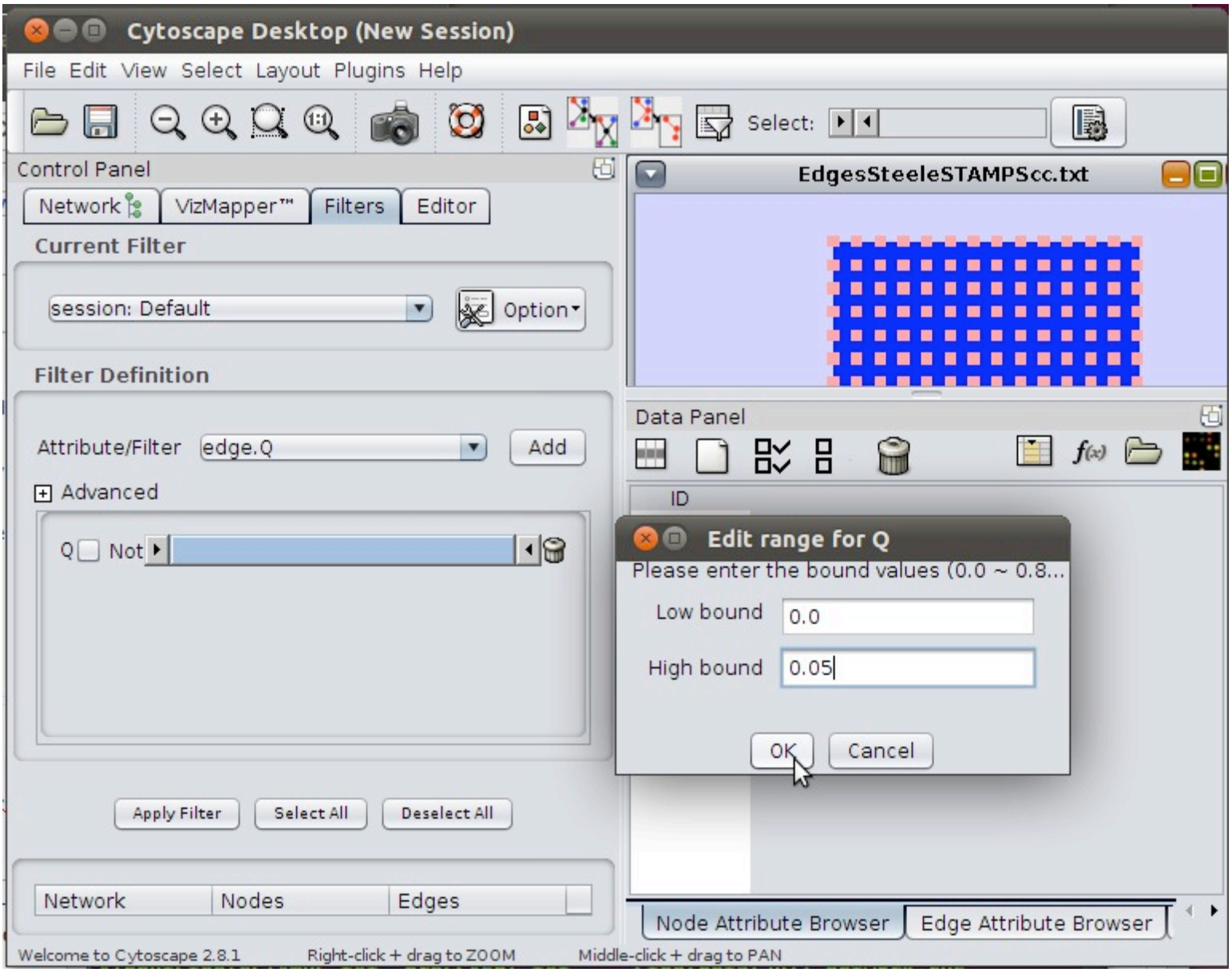
1)Under the *Control Panel*, select the *Filters* tab.

2)In the *Attribute/Filter* menu select **Edge.Q**

3)In the menu below Q, a slider bar will show up. Double click on the bar.

4)Keep Low bound at 0.0 and set the High bound at 0.05.

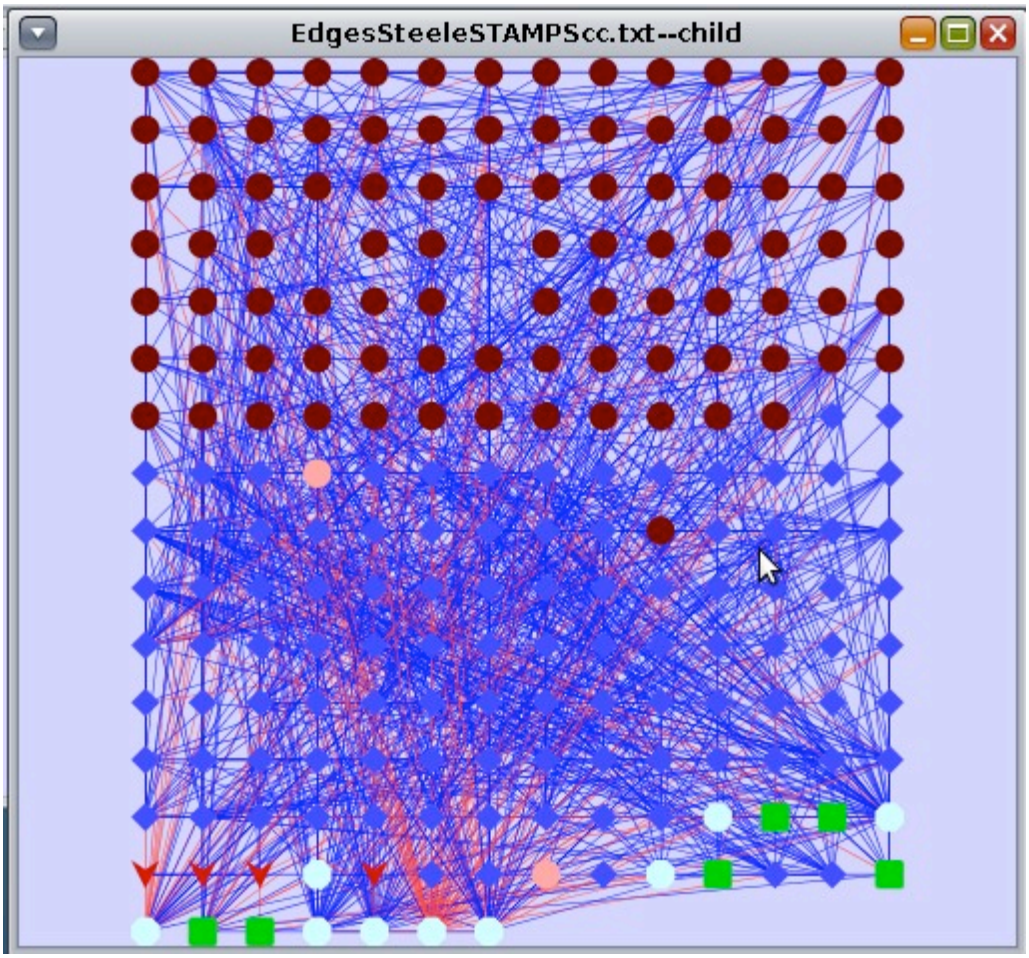
5)Click **OK**



It may be difficult to see, but now only those interactions (edges) with q values better than 0.05 are

Create a New Network from Q<=0.05 values

1)Menu: Select > Nodes > Nodes Connected by Selected Edges
2)Menu: File > New > Networks > From Selected Nodes and Selected Edges



Now we have a file with a more manageable number of edges.

About the Network Panel Window

This is located on the left side. As networks are created, they are listed here.

- Active networks are shown in green, inactive ones are red.
- If you "minimize" a window (yellow button - mac, line - pc), then the network view will remain the same when you return to it.
- If you CLOSE a window, this will *Destroy the view,* meaning that your layout and visual information will not be retained. You can also right-click and select "Destroy view," but again your layout information will not be saved. **Therefore, I recommend just minimizing network windows rather than closing them.**
- To delete a network entirely, right-click and say "Destroy network".

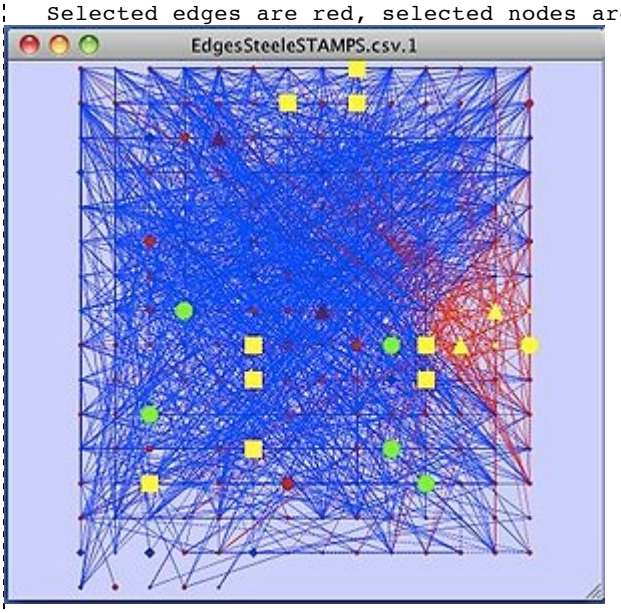
Creating new sub-networks from a few nodes

NOTE: Your networks may look different than those in the tutorial from here on out, depending on what nodes you select.

To show first-order connections from a few manually selected nodes

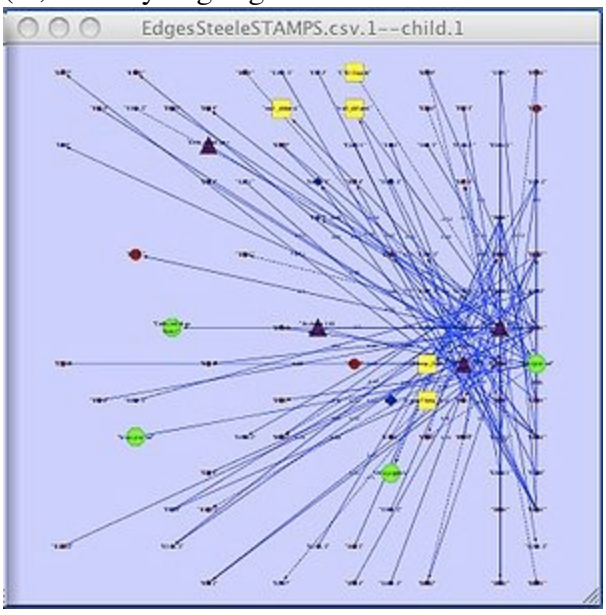
```
1) Navigate back to network window.
```

```
2) Click on magnifying glass with 1:1 to show whole network within the window.
3) Manually select 3-5 nodes individually (shift-click, both Mac and PC) and drag OUTSIDE of the grid s
 Nodes turn yellow, after selection. Remember to continue holding shift until you have moved the nodes.
3) Highlight (drag a box, starting from outside the network area) around the selected nodes and their e
   Selected edges are red, selected nodes are yellow.
```




```
5) Select (in menu bar) Select > Nodes > Nodes connected by selected edges.
6) File > New > Network > From Selected Nodes SELECTED Edges.
```

Result:
New network has the 3-5 nodes you selected and their 1st neighbors, but only with the edges you selected (ie, the only edges go from center node to the 1st neighbor but not from 1st neighbor to other 1st neighbor).

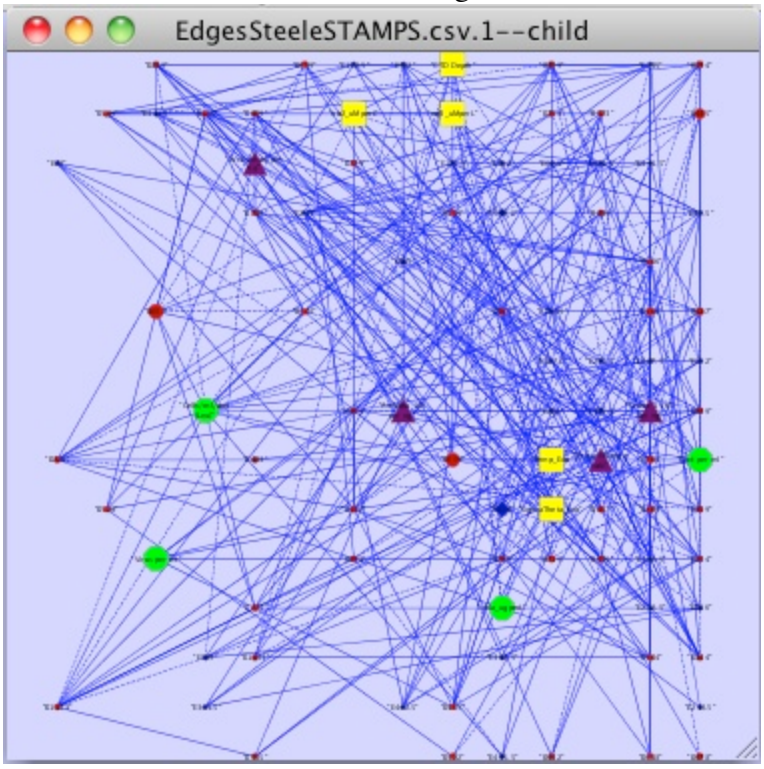


To show all the connections between selected nodes

```
1) Manually select 3-5 nodes (shift-click, both Mac and PC) and drag OUTSIDE of the grid space.
 Nodes turn yellow, after selection. Remember to continue holding shift until you have moved the nodes.
```

```
2) Highlight (drag a box) around those nodes and the edges (selected = red).
3) Select (in menu bar) Select > Nodes > Nodes connected by selected edges.
4) File > New > Network > Select Nodes, ALL edges.
Key difference is choosing ALL edges versus SELECTED edges.
```

Result:
New network should have more edges and show connections between 1st neighbors.



To rename a network
Note: New networks are called .child of parent network. They are listed just below the parent network.
(I highly suggest renaming networks as you make them, or it can get quite confusing 10+ views later.)

```
Right-click or ctrl-click on the network name in "Network" left panel.
A window should open to allow you to rename your network.
```

Creating new sub-networks by searching node and edge attributes

1) To search for specific nodes that share a feature

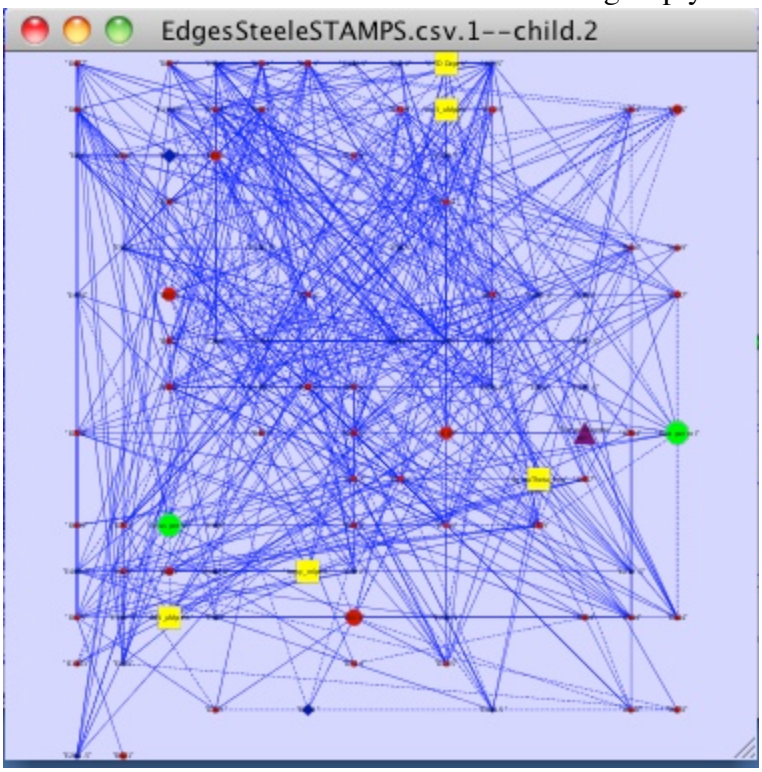
```
1) On left-hand panel, click on original network name to bring that window to the front.
2) Click on Filters tab at top of Control Panel.
   If you don't see "Filters", click on an arrow/triangle to see the other tabs.
3) Click on Option, Create New Filter, Enter name "group"
4) Under Filter Definition, Set Attribute/Filter to "node:Group" and click "Add".
5) Pick your favorite group of bacteria with more than one node from the pull-down menu.
   If a name doesn't appear, start typing a group name and it should appear in the window.
   To view all the options, select all nodes in the network and look at the node attributes (in data pa
   # hits next to Group name indicates how many instances are within your parent network
   Choose a family with more than one hit.
```

```
6) Click Apply (at bottom of panel) which will select all of the nodes within that group.
Example: Rickettsiales for SAR11 group
```

At this point, the information for the selected nodes should appear in the data window. The selected nodes will be in yellow (not to be confused with the nodes you may have colored yellow with the VizMapper).

```
7) Select > Nodes > First Neighbors of Selected Nodes
8) File > New > Network > Selected Nodes, All Edges
```

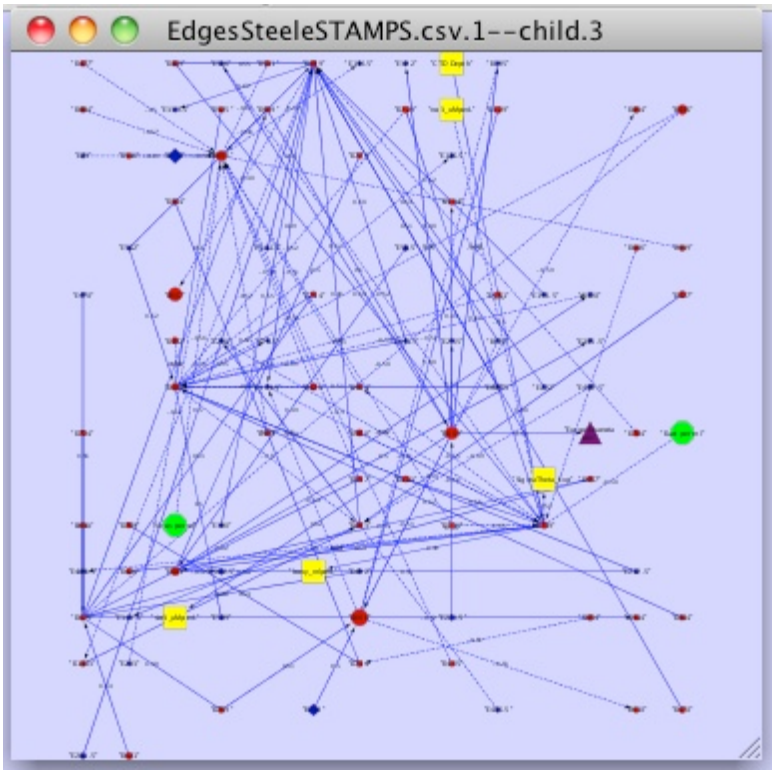
Result from selected nodes, all edges
New Network that is centered on the bacterial group you chose and all connections between the nodes.



2) To Select Nodes and Select Edges for Network

```
1) Repeat Steps 1-5 on the original network to select a new group of bacteria.
2) Select Edges > Select Adjacent Edges. Some of the edges should now appear red.
3) Select > Nodes > Connected by selected edges.
4) File > New > Network > Selected Nodes, Selected Edges
```

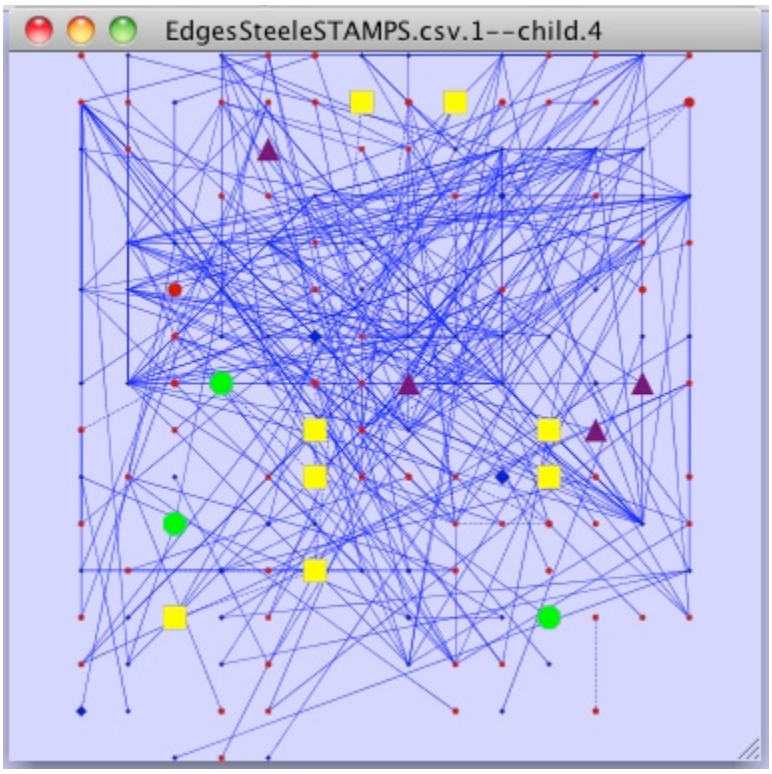
Result
Network shows only connections directly to SAR11, but not the connections between the 1st neighbors. There should be less edges than the previous network.



3) To Select Specific Edges by Attributes
Let's look just at strong LSA connections LSA >0.6 or LSA < -0.6

```
Navigate back to original parent network.
1) Click "Option", and "Create a new filter"
2) Select edge.LS2. Click Add.
3) Adjust sliders so that it contains -0.6 (low bound) to 0.6 (high bound).
(or Double-click on slider and enter numbers)
4) Select "Not".
5) Select Apply Filter at bottom of panel.
6) Select Select > Nodes > Nodes connected by Selected Edges
7) File > New > Network > From Selected Nodes, Selected Edges
```

Result
Only nodes and edges connected by strong correlations are shown. Number of connections should be drastically reduced from original parent network.

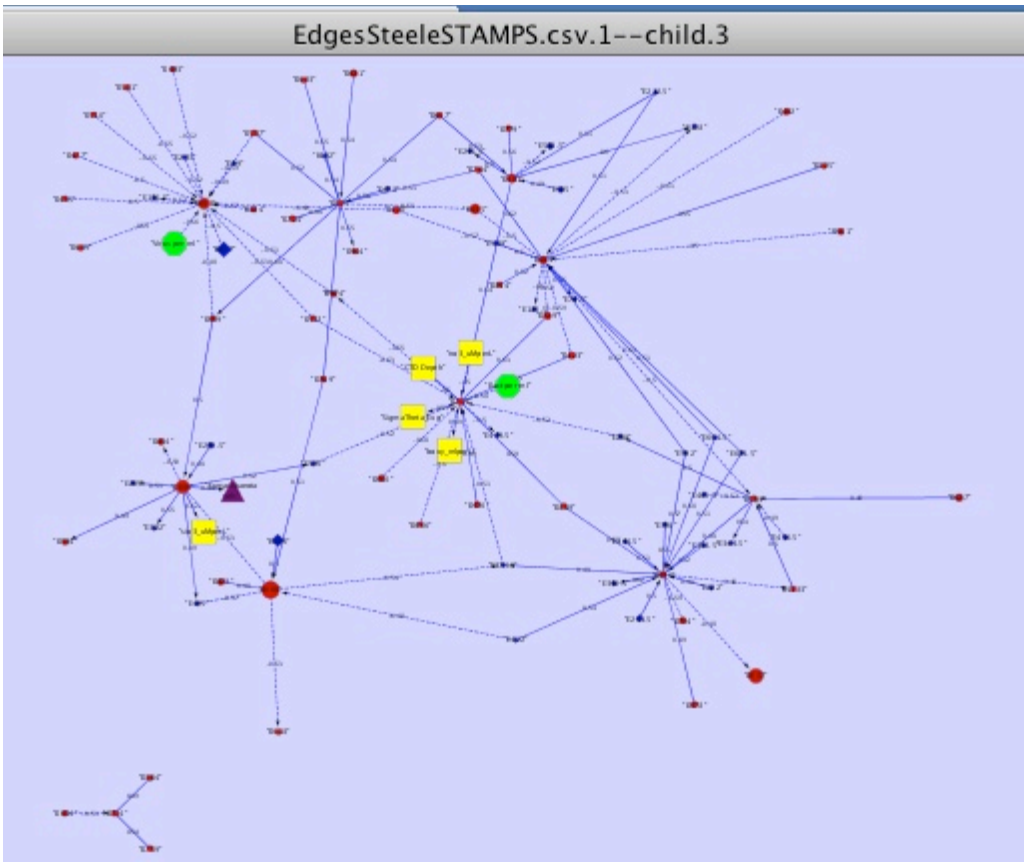


Creating Network Layouts (aka Network Views)

For each (or a few) of the networks you have created

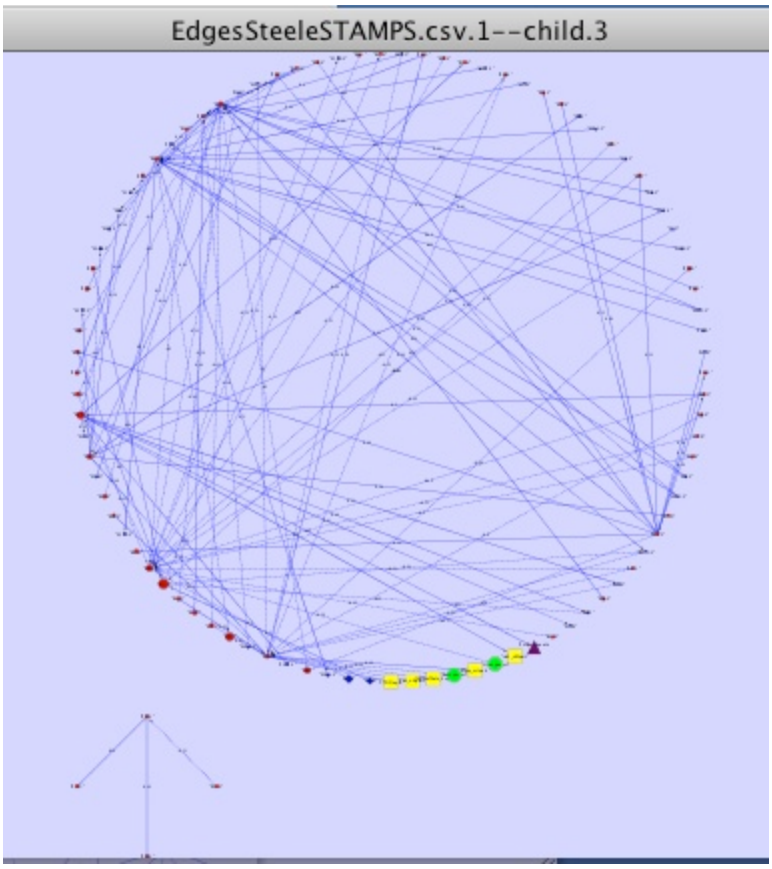
```
Layout > yFiles > Organic

Layout > Cytoscape Layouts > Pick One!
```

Results
The appearance of the network should change, and no longer show the nodes in a square grid.
Organic



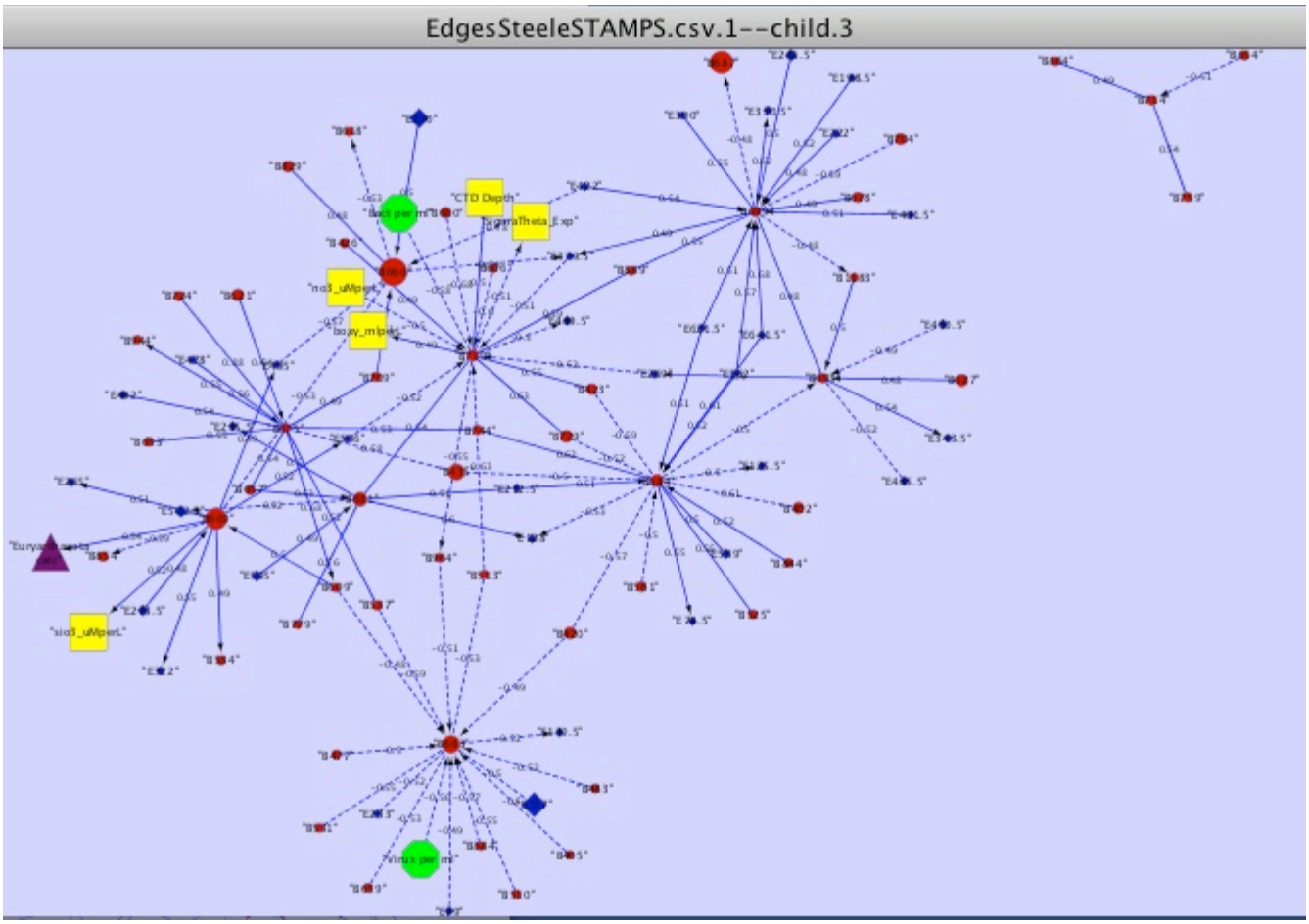
Circle

Note: I chose Attribute Circle Layout. The four nodes in the lower left (that look like an arrow) were connected to a SAR11 OTU, but not any of other OTUs others, and so form their own cluster.

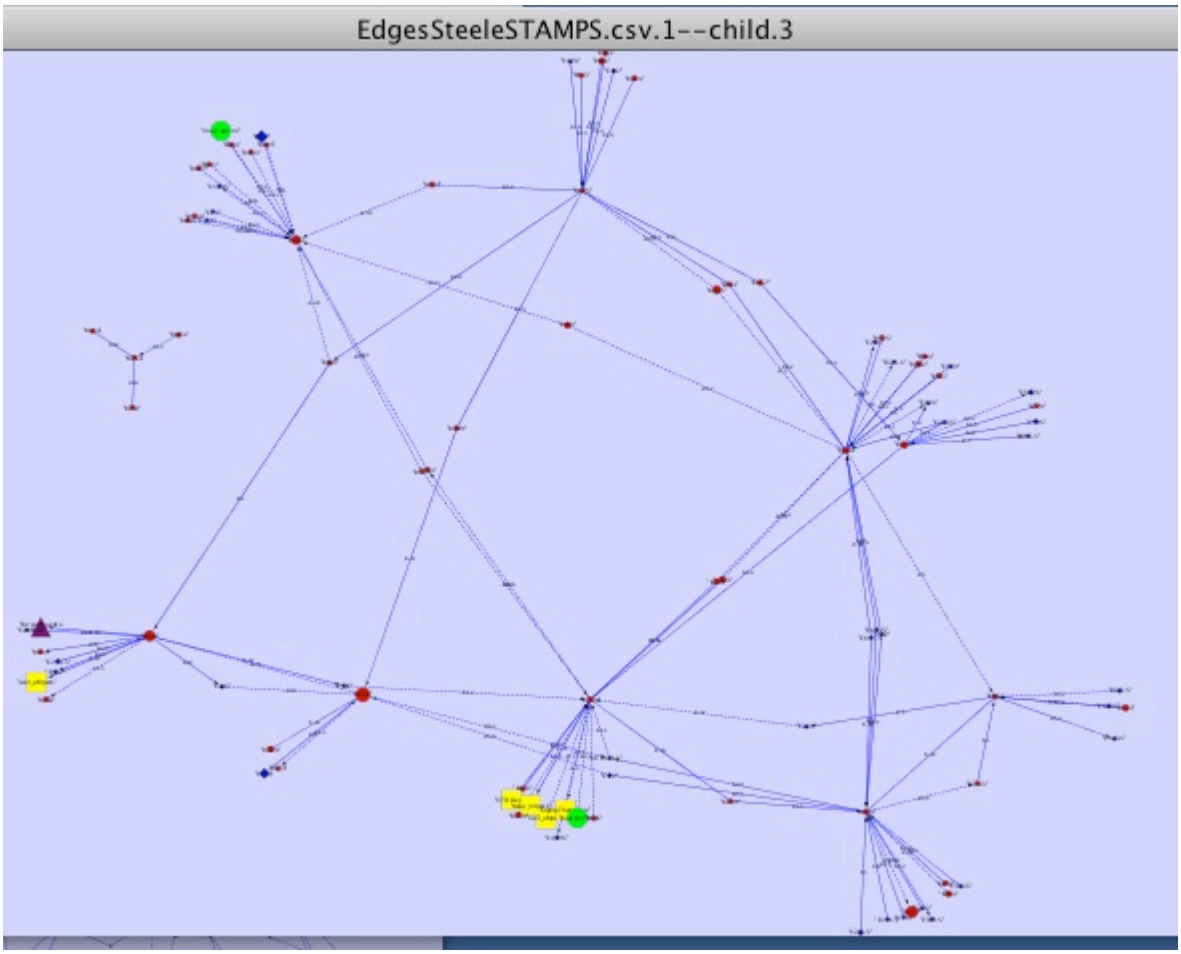


Other Types We Like

Force-Directed - unweighted



Spring-Embedded



Feel free to experiment with the others. Nodes and edges can be highlighted and moved around the screen (at which point, edge lengths no longer have any significance).

More information on what these different layouts illustrate can be found here:
http://opentutorials.cgl.ucsf.edu/index.php/Tutorial:Introduction_to_Cytoscape#Laying_Out_Your_Network

Rotate, Scale, Export Networks

```
In menubar, View > Show Tool Panel
```

Tool Panel will appear in lower left area of window.

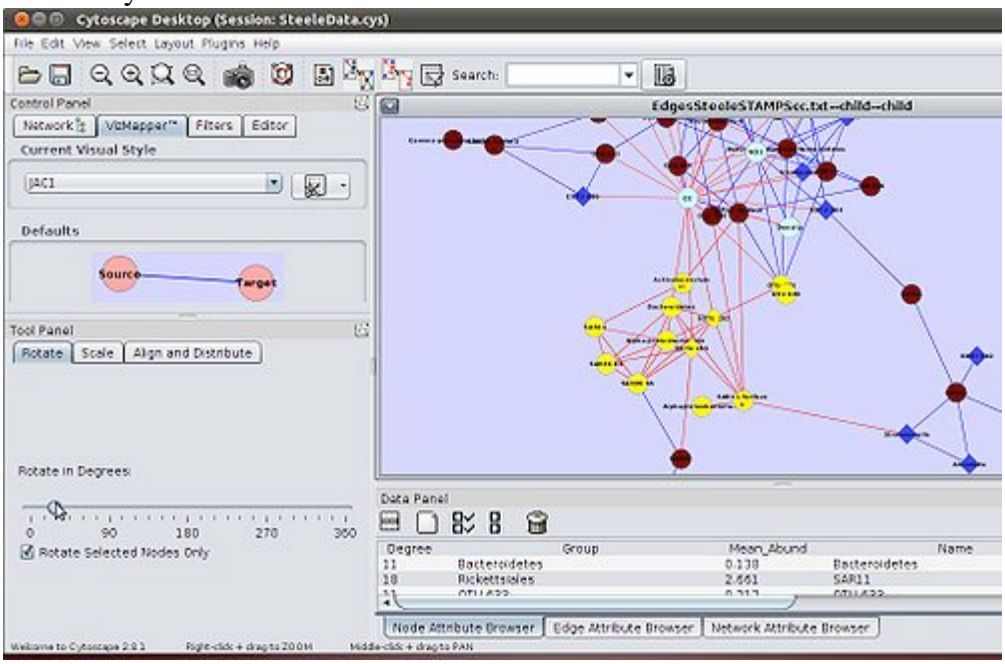
Rotate

```
Click on Rotate.
Slide from 0-180+ degrees.
This will rotate ALL nodes.
```

```
Select some nodes.
```

```
Click "Selected Nodes only
Move slider
```

Note: only some of the nodes rotate.



Scale

```
Click on Scale
Move sliders and play with the other options.
```

Align and Distribute

I don't use this often, but it will move the nodes to various parts of the display windows.

Export your Network as an Image

```
File > Export > Current Network View as Graphics
Choose to select where you are saving the graphics.
PDF is usually fine.  Click OK
*You may need to adjust %magnification and scales to create human-viewable or reasonably-sized files bu
depend on size of your network view and number of nodes/edges.
```

Don't forget to save your work!!

```
File > Save as
```

Congratulations! You made it to the end of the basic tutorial! (http://www.youtube.com/watch?v=Fl4L4M8m4d0)

Feel free to give us (Cheryl, Jacob, or Jed) feedback on this tutorial, it's very much still in development. :)

Answer a Question (aka Play with Cytoscape or Bonus Round)

1.Create a network that might indicate differentiation of SAR11 ecotypes.

   a. Do SAR11 OTUs correlate with abiotic parameters?
   c. What is the strongest correlation (highest LS value) between a SAR11 bacteria and a Eukaryote? What eukaryote is this?

2.Look at all edges with LS values greater than +0.8 or less than -0.8.

   a. Do you notice any clusters that stand out?
   b. What kinds of nodes appear to be in these highly connected clusters?

3. What might prey on my (insert favorite organism here)?
4. Eukaryotic OTUS - more influenced by other eukaryotic or bacterial OTUs?

Useful Links

Local Similarity Analysis Tutorial
(https://stamps.mbl.edu/wiki/index.php/Local_Similarity_Analysis_%28LSA%29_Tutorial)

Papers with network or LSA analysis
Steele et al (in press, advance online pub) Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME http://www.nature.com/ismej/journal/vaop/ncurrent/abs/ismej201124a.html *Recent analysis*

Beman et al (2011) Co-occurrence patterns for abundant marine archaeal and bacterial lineages in the deep chlorophyll maximum of coastal California. ISME
http://www.nature.com/ismej/journal/v5/n7/full/ismej2010204a.html
'Note: LSA only'

JA Fuhrman (2009) Microbial community structure and its functional implications 459, 193-199
http://www.nature.com/nature/journal/v459/n7244/full/nature08058.html

Fuhrman, JA and JA Steele (2008) Community structure of marine bacterioplankton: patterns, networks, and relationships to function AME 53:69-81 http://www.int-res.com/abstracts/ame/v53/n1/p69-81/

Ruan et al (2006) Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. 22 (20): 2532-2538. doi: 10.1093/bioinformatics/btl417
http://bioinformatics.oxfordjournals.org/content/22/20/2532.abstract *Original Description of LSA*

Other Examples of Network Analysis
Chaffron et al., (2010) A global network of coexisting microbes from environmental and whole-genome sequence data. 20: 947-959 http://genome.cshlp.org/content/20/7/947.full

Paver and Kent (2010) Temporal Patterns in Glycolate-Utilizing Bacterial Community Composition Correlate

with Phytoplankton Population Dynamics in Humic Lakes *Microbial Ecology* Volume 60, Number 2, 406-418. http://www.springerlink.com/content/907635058t563q53/

Shade et al., (2010) Differential bacterial dynamics promote emergent community robustness to lake mixing: an epilimnion to hypolimnion transplant experiment. Environmental Microbiology Volume 12, Issue 2, pages 455–466, http://onlinelibrary.wiley.com/doi/10.1111/j.1462-2920.2009.02087.x/full *Recent analysis from a lake environment*

Additional tutorials and manuals

Official Cytoscape Tutorials (http://opentutorials.cgl.ucsf.edu/index.php/Portal:Cytoscape)
Nature Protocols article from creators of Cytoscape (http://www.nature.com/nprot/journal/v2/n10/abs/nprot.2007.324.html)
Make an OTU-Sample network in Qiime (http://qiime.sourceforge.net/scripts/make_otu_network.html) and how to import that data into Cytoscape (http://qiime.sourceforge.net/scripts/cytoscape_usage.html)

NetworkAnalyzer is a Java plugin for Cytoscape and computes specific parameters describing the network topology. [1] (http://med.bioinf.mpi-inf.mpg.de/netanalyzer/)